

Skepsis on the scenario of Biological Evolution provided by stochastic models

Thomas Oikonomou*

*Institute of Physical Chemistry,
National Center for Scientific Research "Demokritos",
15310 Athens, Greece*

*School of Medicine Department of Biological Chemistry,
University of Athens, Goudi,
11527 Athens, Greece*

(Dated: March 25, 2008)

Abstract

Stochastic models, based on random processes, may lead to power law distributions, which provide long range correlations. The observation of power law behavior and the presence of long range correlations in biological systems has been demonstrated in various studies. The combination of the two just mentioned results, theoretical and experimental, supports strongly the scenario of biological evolution across different organisms. In the current Letter we explore in a general way, using the algebra of Nonextensive Statistics introduced by Tsallis and coworkers, if the processes which are described by a class of stochastic models are really random and discuss the results with regard to a possible biological evolution.

PACS numbers: 02.50.-r; 87.23.Kg; 87.10.-e

Keywords: stochastic models; biological evolution; q -Algebra; long range correlations

*Electronic address: thoikonomou@chem.demokritos.gr

In the last years a variety of systems in diverse scientific fields has been explored presenting power law distributions like in biophysics [1], neurophysics [2], economics [3], turbulence [4, 5], urban agglomerations [6] and many others. Because of the ubiquity of power laws many stochastic models have been studied [7, 9], which are based on random processes and lead to the desired power law distributions. These results have a great impact on our point of view about the rules that dominate in nature. Singularly, in biology the success of deriving power laws through the consideration of random processes supports strongly the scenario of the biological evolution across different species. These models do not give explicit description of the transition between different organisms during the evolution but they can form well, under consideration of biological processes like mutation, duplication, etc., symbolic sequences which correspond to polynucleotide DNA chains. For example in Ref. [7] Provata used a random aggregation model to describe a possible formation of such a polynucleotide DNA chain. The model is based on the probability formula

$$P(s, t + \Delta t) = \sum_{n=1}^N P(I) \binom{N}{n} \left(\frac{1}{N}\right)^n \left(1 - \frac{1}{N}\right)^{N-n} \prod_{j=1}^n P(s_j, t) \Big|_{\sum_{j=1}^n s_j + I = s} \quad (1)$$

introduced by Takayasu and collaborators [10, 11], which describes the probability of finding a macromolecule of size s at the time $t + \Delta t$. The analytical expression of Eq. (1) is computed in the transformed Fourier space. A different stochastic approach is considered in Ref. [9] by Messer, Arndt and Lässig, who constructed a master equation for the joint probabilities $P_{\text{eq}}(r, t)$ and $P_{\text{op}}(r, t)$ of finding a *GC* pair and a *AT* pair, respectively, at a distance r

$$\begin{aligned} \frac{d}{dt} P_{\text{eq}}(r, t) = & 2\mu_{\text{eff}} [P_{\text{op}}(r, t) - P_{\text{eq}}(r, t)] \\ & + [r\delta + (r-1)\gamma^+] [P_{\text{op}}(r-1, t) - P_{\text{eq}}(r, t)] \quad \cdot \\ & + r\gamma^- [P_{\text{eq}}(r+1, t) - P_{\text{eq}}(r, t)] \end{aligned} \quad (2)$$

The exchange $P_{\text{eq}}(r, t) \leftrightarrow P_{\text{op}}(r, t)$ in Eq. (2) gives the analogous equation for the probability $P_{\text{op}}(r, t)$. Then, the correlation function $C(r, t) = P_{\text{eq}}(r, t) - P_{\text{op}}(r, t)$ can be computed. The consideration of both Eqs. (1) and (2) leads, under certain assumptions, to power law distributions.

On the other hand, it is well known that the existence of power law distributions indicates the presence of long (or short) range correlations. Yet, this is in contradiction with the statistical assumptions of the stochastic models about random processes. It would mean

that uncorrelated processes can create, in an unexplained way, correlations between them. Even if we assume that in nature such an event (creation of correlations through random processes) can be possible, can we assume the same in mathematics? A stochastic model is a mathematical construction with concrete rules and it can not lead to the existence of correlations if we have not first introduced them in the model. Starting from this point of view we shall explore in the next paragraphs whether a class of stochastic models, where sums appear like in Eqs. (1) and (2), describes really random processes. For this aim we shall use the q -Algebra [12] of Nonextensive Statistics [13]. The results will be discussed in the frame of biology.

Nonextensive Statistics is a possible generalization of the ordinary random statistics, depending on a parameter $q \in [0, 1]$ (the $q \geq 1$ -branche corresponds to different q -statistics [14]). Its structure is based on the following deformed logarithm

$$\ln_q(x) := \frac{x^{1-q} - 1}{1 - q}, \quad (x > 0), \quad (3)$$

with its inverse deformed exponential function

$$e_q(x) := [1 + (1 - q)x]^{\frac{1}{1-q}}, \quad (x > 0). \quad (4)$$

When $q = 1$ the entire generalized statistical structure recovers the ordinary one. One of the main points in Nonextensive Statistics is its successful treatment of correlated probabilities (or processes in general), which are projected on the values of the parameter q .

First, we introduce some elementary notions of the Probability Theory. A *probability space*, which represents our uncertainty regarding an experiment consists of two parts: i) a set of events/outcomes, which is denoted as *sample space* Ω and ii) a real function of the subsets of Ω which is denoted as *probability measure* P . Now, let us follow the next thought. We consider two coins I and II . The sides of each coin have a different symbol, so that the sample spaces Ω_I and Ω_{II} consist of two events, $\Omega_I = \{A, B\}$ and $\Omega_{II} = \{\tilde{A}, \tilde{B}\}$. Then, the *intersection probability* (of both occurring) in each sample space separately is of course equal to zero

$$P(A \cap B) = 0 = P(\tilde{A} \cap \tilde{B}). \quad (5)$$

The *union probability* (of either occurring) in both spaces separately is equal to the unity

$$\left. \begin{aligned} P(A \cup B) &= P(A) + P(B) \\ P(\tilde{A} \cup \tilde{B}) &= P(\tilde{A}) + P(\tilde{B}) \end{aligned} \right\} = 1. \quad (6)$$

Eq. (6) corresponds in physics to the *normalization constraint*. In this probability measure the event probabilities in each sample space are strongly correlated in such a way that their sum is always equal to the unity. We denote these correlations as correlations of Type I. On the other hand, with respect to the values of the probabilities, there are infinite ways to fulfil Eq. (6). In order to capture all these possibilities, or better to attribute them to a cause, we introduce yet another type of correlations, which we call correlations of Type II. These correlations are individual for each event. In our example they could represent the different weights of each side of the coin. When the correlations of Type II vanish, then we have equal event probabilities.

In further, we join the spaces $\Omega_{I,II}$. Now, we have one more type of correlations to consider, which we denote as correlations of Type III. These correlations are between the several combinations of the events. Following these steps, the consideration of more complicate processes or compositions would lead to an increasing number of the correlation types (higher order correlations, Type $>III$). In shake of the simplicity, we assume that there are no correlations of the Types II and III in the new sample space Ω_{III} and because of this, Ω_{III} is a tensor product of both subspaces $\Omega_{III} = \Omega_I \times \Omega_{II} = \{A\tilde{A}, A\tilde{B}, B\tilde{A}, B\tilde{B}\}$. In this case the event probabilities $P(A)$, $P(B)$, $P(\tilde{A})$ and $P(\tilde{B})$ are equal and independent. Then, the intersection probabilities are given by

$$P(A \cap \tilde{A}) = P(A)P(\tilde{A}), \quad (7)$$

$$P(A \cap \tilde{B}) = P(A)P(\tilde{B}), \quad (8)$$

$$P(B \cap \tilde{A}) = P(B)P(\tilde{A}), \quad (9)$$

$$P(B \cap \tilde{B}) = P(B)P(\tilde{B}), \quad (10)$$

and because of the union probability

$$\begin{aligned} P(A\tilde{A} \cup A\tilde{B} \cup B\tilde{A} \cup B\tilde{B}) &= P(A)P(\tilde{A}) + P(A)P(\tilde{B}) + \\ &P(B)P(\tilde{A}) + P(B)P(\tilde{B}) = 1, \end{aligned} \quad (11)$$

their values are equal to $1/4$. Here, there is a very important point to stress. Namely, the feature of the probability (in)dependence is meaningful in the frame of an appropriate probability measure. As we can see in Eqs. (7) - (10), the intersection probabilities of Ω_{III} are computed based on the independence of the event probabilities between the two subspaces Ω_I and Ω_{II} . Yet, the union probability of Ω_{III} correlates strongly the intersection

probabilities. Furthermore, it becomes evident that when we refer to correlations we have to be very concrete about their Type. To give a more general character to the intersection probabilities for the further purposes of this Letter we shall call them *Blocks of probabilities*.

In Ref. [14] the present author demonstrated the importance of the distinction between the diverse types of correlations. The existence of two different types of correlations within Nonextensive Statistics has been introduced, the *inner* and the *outer*, described by a set of parameters $\mathcal{R} = \{R_i\}_{i=1,\dots,n}$ and $\mathcal{Q} = \{Q_i\}_{i=1,\dots,n'}$ respectively, in order to distinguish between the three generalized entropic structures that can be created, based on any generalized logarithmic function. The inner and outer correlations in this case should not be necessarily identified with ones in our example above. Their notion must be every time adjusted to the features of a problem under consideration. For the one parametric ($\mathcal{Q} = \{Q_1\} = q$) generalized logarithm (3) with its inverse function (4) these three structures correspond to the Tsallis, Rényi and Nonextensive Gaussian entropy. We shall keep the notation of the correlation-sets \mathcal{R} and \mathcal{Q} throughout the paper.

For an arbitrary deformed logarithmic and exponential function we define the generalized product between two elements $x, y \in \mathbb{R}_+$ as [14]

$$x \otimes_{\mathcal{Q}} y := e_{\mathcal{Q}}(\ln_{\mathcal{Q}}(x) + \ln_{\mathcal{Q}}(y)). \quad (12)$$

For $\mathcal{Q} \rightarrow \mathcal{Q}_0$ we obtain the ordinary multiplication since $e_{\mathcal{Q}_0}(\ln_{\mathcal{Q}_0}(x) + \ln_{\mathcal{Q}_0}(y)) = e(\ln(x) + \ln(y)) = x \times y$. Then, for the \mathcal{Q} -multiplication of m elements we have

$$\prod_{k=1}^m_{\otimes_{\mathcal{Q}}} x_k = e_{\mathcal{Q}}\left(\sum_{k=1}^m \ln_{\mathcal{Q}}(x_k)\right), \quad (x_k > 0). \quad (13)$$

Using the deformed functions in Eqs. (3) and (4), Eq. (13) takes the explicit form

$$\prod_{k=1}^m_{\otimes_q} x_k = \left[\sum_{k=1}^m x_k^{1-q} - (m-1) \right]^{\frac{1}{1-q}}. \quad (14)$$

If we express the normalization constraint of W intersection probabilities P_j through the q -algebra we can easily verify the relation

$$\sum_{j=1}^W P_j = W \otimes_q \prod_{j=1}^W_{\otimes_q} P_j \Big|_{q=0} = 1. \quad (15)$$

The value $q = 0$ in q -Statistics corresponds to long range correlations. This result reproduce correctly the correlations of Type I behind the normalization constraint, which relate all

probabilities in a very specific way. Since generalized statistics gives us proper results and does not lead to inconsistencies we shall use it in the next paragraph in order to complete the aim of this Letter.

We consider the statistical processes of a system and define the probabilities under which these processes take place and the way they interact (correlated or uncorrelated). An example for such a system under consideration could be a DNA chain and a possible statistical process could be the mutation of a nucleotide. In connection to our example with the coins, all these interacting probabilities define the respective Blocks of probabilities for the current approach and are denoted as $\{A_k^{(\mathcal{R})}(P_j)\}_{k=1,\dots,m}$. The internal \mathcal{R} -correlations characterize the different types of correlations within a Block. These correlations are canceled for certain values $\mathcal{R} \rightarrow \mathcal{R}_0$. Now, we want to see how the several k -Blocks interact with each other. These interactions are described by the external correlations \mathcal{Q} . Then, we introduce the probability functional $B_{\{\mathcal{Q}, \mathcal{R}\}}(P_j)$ which is equal to the entire internal and external interactions. It could be a derivative of the probabilities P_j in $A_k^{(\mathcal{R})}(P_j)$, as in Eq. (2), or simply a new probability, as in Eq. (1), etc. The elimination of the external correlations is defined again in a certain limit $\mathcal{Q} \rightarrow \mathcal{Q}_0$. Then, the entire process is given by an expression of the form

$$B_{\{\mathcal{Q}, \mathcal{R}\}}(P_j) = m \otimes_{\mathcal{Q}} \prod_{k=1}^m A_k^{(\mathcal{R})}(P_j), \quad (16)$$

with $B_{\{\mathcal{Q}, \mathcal{R}\}}(P_j) > 0$. We stress that the Eq. (16) does not describe exact processes. It is a general expression of a set of probability interactions. The \mathcal{Q} -multiplication with the constant m , does not have a specific meaning but is useful for our purpose. We demand that the stochastic processes inside of each Block are independent ($\mathcal{R} \rightarrow \mathcal{R}_0$), so that they are multiplicatively connected through the ordinary multiplication. Using Eq. (14), the relation in Eq. (16) can be written as

$$B_{\{q, \mathcal{R}_0\}}(P_j) = \left[m(q) + \sum_{k=1}^m \left[A_k^{(\mathcal{R}_0)}(P_j) \right]^{1-q} \right]^{\frac{1}{1-q}}. \quad (17)$$

with $m(q) := m^{1-q}(1 - m^q)$. In further, we set $q = 0$ as we have done in Eq. (15). Then, from Eq. (17) we obtain

$$B_{\{q=0, \mathcal{R}_0\}}(P_j) = \sum_{k=1}^m \left[A_k^{(\mathcal{R}_0)}(P_j) \right]. \quad (18)$$

Eq. (18) is our main result. Namely, when the probability measure $B_{\{q, \mathcal{R}_0\}}(P_j)$ is equal to the sum ($q = 0$) of the probability Blocks, even if the internal processes inside a Block are independent, then this measure describes long range correlations. Accordingly, every stochastic model, which is based on such a sum, gives a possible creation of sequences through a very concrete dynamic, where the processes of the several probability Blocks are long range correlated (synchronized). Comparing Eqs. (1) and (2) with Eq. (18), we see that the two earlier equations are specific expressions of the latter one.

If we accept that biological evolution, as described by Eq. (18), has taken place, then the transition between different species during the evolution is attributed to the combination of two different kinds of modifications in the \mathcal{R} -correlations in each Block. The first kind of modifications do not affect the structure of the given dynamic. This case would correspond to the creation of different organisms of the same class. The second kind of modifications does affect the structure of the given dynamic. This case would correspond to the creation of different classes of organisms. However, the existence of long range correlations has the effect that a modification of a single probability causes simultaneously changes in all involved terms. This has as consequence that every new created organism does not have the time to be adapted to its environment, so that it is able to survive. On the other hand, the adaption to the environment is the basis of the natural selection and the biological evolution. As we can see, the introduction of long range correlations gives a problematic to the scenario of biological evolution, as has been assumed up to now.

Summarizing, we have shown that the dependence or independence between probabilities is associated to an appropriate probability measure. The sum of probabilities or of a Block of probabilities corresponds to long range interactions between the participating terms. Accordingly, any stochastic model in which such a sum appears, correlates in a very specific way, with long range features, the respective processes. The result is a non-exponential probability distribution function. Beside the clarification of the appropriate probability measure, we need to clarify the Type (or Types) of the correlations as well. There is a variety of different types of correlations in a process, whose confusion may lead to incorrect results. The more complicated a process is, the more different types of correlations are involved in this process. The presence of correlations creates difficulties considering the cornerstone of the

natural selection in biology, which is the adjustment of an organism to its environment.

- [1] Th. Oikonomou, A. Provata, U. Tirnakli, Physica A **387**, 2653 (2008).
- [2] E. Novikov, A. Novikov, D. Shannahoff-Khalsa, B. Schwartz, J. Wright, Phys. Rev. E **56**, R2387 (1997).
- [3] Tsallis, C., Anteneodo, C., L. Borland and R. Osorio, Physica A **324**, 89 (2003).
- [4] Beck, C., Lewis, G.S., Swinney, H.L., Phys. Rev. E **63**, 035303 (2001).
- [5] Arimitsu, N., Arimitsu, T., Europhys. Lett. **60**, 60 (2002).
- [6] Malacarne, L.C., R.S. Mendes, Lenzi, E.K., Phys. Rev. E **65**, 017106 (2002).
- [7] A. Provata, Physica A **264**, 570 (1999).
- [8] A.M.C. de Souza & C. Anteneodo, Biophysical Journal **69**, 1708 (1995).
- [9] P.W. Messer, P.F. Arndt & M. Lässig, Phys. Rev. Lett. **94**, 138103 (2005).
- [10] H. Takayasu, M. Takayasu, A. Provata, G. Huber, J. Stat. Phys. **65**, 725 (1991).
- [11] H. Takayasu, Phys. Rev. Lett. **63**, 2563 (1989).
- [12] E.P. Borges, Physica A **340**, 95 (2004).
- [13] J.P. Boon, C. Tsallis, Europhysics News **36**, 6 (2005).
- [14] Th. Oikonomou, Physica A **386**, 119 (2007).